2Б

УДК 62-50

П.І. Бідюк, О.М. Терентьєв, М.М. Коновалюк
Інститут прикладного системного аналізу НТУУ «КПІ», м. Київ, Україна
pbidyuke@gmail.com, PhDMHK@mail.ru, konovalyuk@gmail.com

Байєсівські мережі в технологіях інтелектуального аналізу даних

У статті запропонований огляд методів побудови (навчання) структури мереж Байєса. Показано, що на сьогодні існує безліч методів структурного навчання МБ та критеріїв оптимізації, які можна використати при їх побудові. Тому вибір методу навчання структури мережі повинен ґрунтуватись на докладному поглибленому аналізі задачі, яка розв'язується за допомогою мережі, та можливості отримання достовірних експертних і статистичних даних.

1 Інтелектуальний аналіз даних

Об'єм інформації у світі з кожним роком стрімко збільшується, відбувається перевантаження інформацією. Ця інформаційна лавина надходить з науки, бізнесу, Інтернету та інших джерел. Ускладнюється робота аналітика, який повинен проаналізувати великі масиви інформації при розв'язуванні поставленої задачі. Він розв'язує її, виходячи зі своїх знань і досвіду. Але знання є не лише у людини, вони містяться також у накопичених даних, які необхідно аналізувати. Такі знання часто називають «прихованими», оскільки вони вимагають для зберігання гігабайтів і терабайтів інформації, які людина не в змозі дослідити самостійно.

Очевидно, що для виявлення прихованих знань потрібно застосовувати спеціальні методи автоматичного аналізу даних, за допомогою яких доводиться добувати знання із величезного об'єму інформації. «Через велику кількість інформації дуже мала її частина буде коли-небудь побачена людським оком. Наша єдина надія – зрозуміти та знайти щось корисне у цьому океані інформації – це широке застосування методів Data Mining», – відзначив один із засновників цього напряму Григорій П'ятецький-Шапіро (Gregory Piatetsky-Shapiro) [1].

Існують різні означення Data Mining. Але вони збігаються у головному, оскільки мають чотири основні ознаки, які наявні у кращому означенні технології Data Mining, запропонованому у 1996 р. Г. П'ятецьким-Шапіро: «Data Mining – дослідження та виявлення "машиною" (алгоритмами, засобами штучного інтелекту) в сирих даних прихованих, раніше невідомих, нетривіальних, практично корисних та доступних для інтерпретації людиною знань, необхідних для прийняття рішень у різних сферах людської діяльності» [2].

Суть і ціль технології Data Mining полягає у пошуку неочевидних, об'єктивних і корисних на практиці закономірностей у великих обсягах даних. В основу сучасної технології Data Mining покладена концепція шаблонів (патернів), які відображають фрагменти багатоаспектних взаємовідносин в даних. Ці фрагменти представляють собою закономірності, властиві підвибіркам даних, які можуть бути компактно виражені у зрозумілій людині формі. Пошук шаблонів здійснюється методами, не обмеженими рамками апріорних припущень стосовно структури вибірки та виду розподілу значень змінних, що аналізуються.

Поняття Data Mining з'явилося у 1989 році, але високу популярність у сучасному трактуванні набуло приблизно у першій половині 1990-х років. Data Mining широко використовується у багатьох галузях, для яких характерне використання великих об'ємів даних: в науці, торгівлі, телекомунікаційній сфері, банківський справі, промисловому виробництві та інших галузях, де виникає задача автоматичного аналізу даних і прийняття рішень на його основі. Завдяки мережі Інтернет Data Mining використовується кожний день користувачами пошукових систем на просторах Інтернету.

Data Mining лежить на перетині декількох наук, основні з яких – це системи баз даних, статистика та штучний інтелект. До *методів та алгоритмів Data Mining* відносять такі: штучні нейронні мережі, дерева рішень, символьні правила, методи найближчого сусіда і k-найближчого сусіда, метод опорних векторів, байєсівські мережі, лінійну регресію, кореляційно-регресійний аналіз; ієрархічні методи кластерного аналізу, неієрархічні методи кластерного аналізу, у тому числі алгоритми k-середніх і k-медіани; методи пошуку асоціативних правил, наприклад, алгоритм Apriori; метод обмеженого перебору, еволюційне програмування і генетичні алгоритми, різноманітні методи візуалізації даних та інші методи.

До задач Data Mining відносять: класифікацію, кластеризацію, асоціацію, послідовну асоціацію або просто послідовність, прогнозування, визначення відхилень або викидів, оцінювання параметрів і станів, аналіз зв'язків, візуалізацію, підбиття підсумків.

2 Байєсівська мережа – інструмент інтелектуального аналізу даних

Інтелектуальний аналіз даних (*ІАД*) – мультидисциплінарна область, що виникла та розвивається на базі таких наук, як прикладна статистика, розпізнавання образів, штучний інтелект, теорія баз даних та ін. (рис. 1).

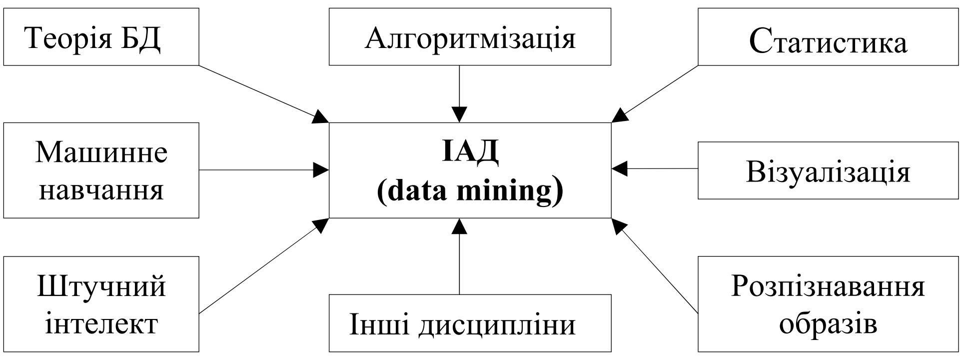


Рисунок 1 – ІАД як мультидисциплінарна область

Перед використанням технологій ІАД необхідно ретельно проаналізувати можливі проблеми, обмеження та критичні питання, які зв'язані з нею, а також зрозуміти те, чого ця технологія не може дати. Очевидно, що технологія ІАД не може дати відповіді на ті питання, які не були задані. Вона не може замінити аналітика, а всього лише дає йому потужний інструмент для полегшення і підвищення якості його роботи.

Оскільки технологія ІАД – мультидисциплінарна область, то для розробки програмного забезпечення, що включає ІАД, необхідно задіяти фахівців з різних галузей, а також забезпечити їх високоякісну взаємодію. Неможливо видобувати корисну інформацію без розуміння суті даних. Використання ІАД має бути нерозривно пов'язаним із підвищенням кваліфікації користувача. Більшість інструментів інтелектуального аналізу даних ґрунтується на двох технологіях: машинне навчання (machine learning) і візуалізація (візуальне подання інформації). Ці дві технології якраз і поєднують у собі ***байєсівські мережі*** **(**БМ**).** Це відносно молодий напрям розвитку науки, що з'явився на стику теорії ймовірностей і теорії графів (рис. 2).

БМ – це графи із деякими характерними властивостями. Ідея впровадження БМ полягає у представленні причинно-наслідкових зв'язків, характерних для процесу у вигляді графа.

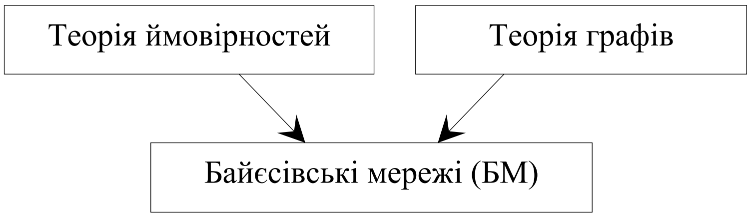


Рисунок 2 – БМ на стику двох наук

Томас Байєс одним з перших зацікавився ймовірністю настання подій у майбутньому, ґрунтуючись на інформації про минулі випробування. Саме теорема Байєса пов'язує апріорні та апостеріорні ймовірності причин після спостереження за наслідками. До впровадження терміна «байєсівська мережа» Джуді Перл застосовував БМ під назвою каузальних мереж (causal network), тобто мережі з причинно-наслідковими зв'язками. Байєсівськими вони стали завдяки застосуванню в каузальних мережах *теореми Байєса*.

Теорема Байєса. Нехай $H_1, H_2,..., H_n$ – попарно несумісні події і їх сума збігається з усім вибірковим простором подій. Тоді для будь-якої випадкової події *X*, що може з'явитися лише за умови появи однієї з подій $H_1, H_2,..., H_n$, і такої, що $P(X) \neq 0$, виконуються рівності:

$$p(H_k \mid X) = \frac{p(X \mid H_k) \cdot p(H_k)}{\sum_{i=1}^{n} p(X \mid H_i) \cdot p(H_i)}, \quad k = \overline{1, n} \tag{1}$$

В (1) H_k означає будь-яку гіпотезу з n можливих. Ймовірності $p(X \mid H_k)$ задаються експертами *апріорно* або розраховуються за навчальними даним. Тобто їх можна розглядати як відповідь на запитання: «*Якою буде ймовірність деякого виміру, якщо відомо, яка гіпотеза була реалізована*?». Ймовірності $p(X \mid H_k)$ є дуже корисними, тому що, як правило, легше знайти ймовірність послідовності подій типу *причина-наслідок*, ніж навпаки. Значення $p(H_k)$ називають *апріорними ймовірностями*, вони визначають початкові ймовірності для всіх гіпотез. Потужність байєсівського методу полягає у тому, що апріорні ймовірності можна уточнювати (оновлювати) відповідно до фактичних реалій перебігу процесу, що досліджується. Це дозволяє уточнювати ймовірності подій при надходженні додаткової інформації.

2.1 Переваги застосування байєсівських мереж

У рамках технології ІАД головна цінність БМ полягає у їх здатності виявляти невідомі та нетривіальні зв'язки між факторами, про які іноді самі експерти у відповідній предметній області не мають уявлення. Байєсівські мережі знаходять своє практичне застосування у таких сферах, як медицина, фінанси та економіка, комп'ютери і системне програмне забезпечення, обробка зображень та відео, військова справа, космічні польоти та дослідження, а також багато інших.

На відміну від інших методів ІАД, застосування байєсівських мереж до аналізу процесів різної природи, діяльності людини та функціонування технічних систем дозволяє враховувати та використовувати будь-які вхідні дані у вигляді експертних оцінок і статистичної інформації. У свою чергу, змінні можуть бути дискретними і неперервними, а характер їх надходження при аналізі та прийнятті рішення може бути в режимі реального часу і у вигляді статичних масивів інформації і баз даних. При цьому завдяки використанню представлення взаємодії між факторами процесу у вигляді причинно-наслідкових зв'язків у мережі досягається максимально високий рівень візуалізації та чітке розуміння суті взаємодії факторів процесу між собою. Іншими перевагами БМ є можливості врахування невизначеностей статистичного, структурного і параметричного характеру, а також формування висновку за допомогою різних методів – наближених і точних. Загалом можна сказати, що *БМ – це високоресурсний метод ймовірнісного моделювання процесів довільної природи з невизначеностями різних типів, який забезпечує можливість достатньо точно описувати їх функціонування, оцінювати прогнози та будувати системи управління.*

2.2 Математичний опис байєсівської мережі

БМ представляє собою пару $<G, B>$, у якій перша компонента G – це спрямований нециклічний граф, що відповідає змінним процесу, що досліджується, і записується у вигляді причинно-наслідкової мережі. Друга компонента пари B – це множина параметрів, що визначають мережу. Ця компонента містить параметри $\Theta_{X^{(i)}|pa(X^{(i)})} = P\left(X^{(i)} \middle| pa\left(X^{(i)}\right)\right)$ для кожного можливого значення $x^{(i)} \in X^{(i)}$ та $pa(X^{(i)}) \in$ $\in Pa(X^{(i)})$, де $Pa(X^{(i)})$ позначає набір батьків змінної $X^{(i)} \in G$. Кожній змінній $X^{(i)} \in G$ відповідає окрема вершина. Якщо розглядають більше одного графа, то для визначення батьків змінної $X^{(i)}$ в графі G використовують позначення $Pa^G(X^{(i)})$. Повна спільна ймовірність БМ обчислюється за формулою:

$$P_B(X^{(1)}, \ldots, X^{(N)}) = \prod_{i=1}^{N} P_B(X^{(i)} \,|\, Pa(X^{(i)})).$$

З математичної точки зору БМ – це модель подання наявних і відсутніх ймовірнісних залежностей. При цьому зв'язок $A \to B$ є причинним, якщо подія A є причиною виникнення B, тобто коли існує механізм, відповідно до якого значення, прийняте A, впливає на значення, прийняте B. БМ називають причинною (каузальною), якщо всі її зв'язки причинні.

Насправді байєсівська методологія набагато ширша, ніж сімейство засобів маніпулювання з умовними ймовірностями в орієнтованих графах. Вона включає в себе також моделі із симетричними зв'язками (випадкові поля та решітки), моделі

динамічних процесів (ланцюги Маркова), а також широкий клас моделей із прихованими змінними, що дозволяють вирішувати задачі ймовірнісної класифікації, розпізнавання образів та прогнозування. Нові галузі застосування такі: (1) динамічні процеси і динамічне програмування; (2) оптимальне керування стохастичними системами; (3) прийняття рішень в автономних інтелектуальних системах.

2.3 Типи байєсівських мереж

1. *Дискретні БМ* – мережі, у яких змінні вузлів представлені дискретними величинами. Дискретні БМ мають такі властивості:

– кожна вершина представляє собою подію, що описується випадковою величиною, яка може мати кілька станів;

– всі вершини, пов'язані з «батьківськими», визначаються таблицею умовних ймовірностей (ТУЙ) або функцією умовних ймовірностей;

– для вершин без «батьків» ймовірності їх станів є безумовними (маргінальними).

Інакше кажучи, у байєсівських мережах довіри вершини представляють собою випадкові змінні, а дуги – ймовірнісні залежності, які визначаються через таблиці умовних ймовірностей. ТУЙ кожної вершини містить ймовірності станів цієї вершини за умови конкретних значень станів її «батьків».

2. *Динамічні БМ* – мережі, у яких значення вузлів змінюються з часом, тобто це мережа, яка описує стани динамічної системи.

Динамічні БМ ідеально підходять для моделювання процесів, які змінюються у часі. Їх перевага полягає у тому, що вони використовують табличне представлення умовних ймовірностей, що полегшує, наприклад, представлення різних нелінійних явищ. Треба підкреслити, що термін «часова байєсівська мережа» (temporal Bayesian network) краще відображає суть, ніж «динамічна байєсівська мережа» (dynamic Bayesian network), оскільки тут передбачається, що структура моделі не змінюється. Зазвичай параметри моделі не змінюються з часом, але до структури мережі завжди можна додати додаткові приховані вузли для уточнення опису поточного стану процесу.

3. *Неперервні БМ* – мережі, в яких змінні вузлів – це неперервні величини. У багатьох випадках події можуть приймати будь-які стани з деякого діапазону. Тобто змінна X – неперервна випадкова величина, простором можливих станів якої є весь діапазон її допустимих значень $X = \left\{ x \middle| a \le x \le b \right\}$, що містить нескінченну множину точок. У цьому випадку некоректно говорити про ймовірності окремого стану, тому що при їх нескінченно великій кількості вага кожного буде наближатись до нуля. Тому розподіл ймовірностей для неперервної випадкової величини визначається інакше, ніж у дискретному випадку; для їх опису використовують функції розподілу ймовірностей і щільності розподілу ймовірностей. Неперервні БМ використовують для моделювання стохастичних процесів у просторі станів з неперервним часом.

4. *Гібридні БМ* – мережі, які містять вузли з дискретними і неперервними змінними. При використанні БМ, що містять неперервні і дискретні змінні, існує ряд обмежень:

1 – дискретні змінні не можуть мати неперервних батьків;

2 – неперервні змінні повинні мати нормальний закон розподілу, умовний на значеннях батьків;

3 – розподіл неперервної змінної X з дискретними батьками Y та неперервними батьками Z є нормальним:

$$P(X|Y=y,Z=z)=N(\mu_x(\mu_y,\mu_z),\sqrt{\sigma_x(\sqrt{\sigma_y})}),$$

де μ_x,μ_y,μ_z – математичні сподівання, σ_x,σ_y – дисперсії, $\sqrt{\sigma_x},\sqrt{\sigma_y}$ – середньоквадратичні відхилення; μ_x лінійно залежить від неперервних батьків, а σ_x взагалі не залежить від неперервних батьків. Однак μ_x та σ_x залежать від дискретних батьків. Це обмеження гарантує можливість формування точного висновку.

3 Методи оцінювання структури байєсівських мереж

Більшість існуючих методів оцінювання (побудови) структури БМ можна умовно розділити на дві категорії: (1) на основі *оціночних функцій* (search & scoring) та (2) на основі *тесту на умовну незалежність* (dependency analysis). Більшість із існуючих методів зустрічаються з такими проблемами:

1. Наявність упорядкованої множини вершин (УМВ). У більшості методів, особливо розроблених раніше, вважається, що УМВ задана, але при обробці реальних даних це дуже часто не відповідає дійсності.

2. Низька обчислювальна ефективність. Деякі сучасні методи працюють без використання УМВ, а замість неї використовують тест на умовну незалежність (ТУН). Однак в цьому випадку необхідно виконати експоненціальну кількість таких тестів, що призводить до зменшення ефективності роботи методу у зв’язку із значним зростанням об’єму обчислень.

3. Проблема побудови великих БМ. Існують методи, за допомогою яких можна побудувати структуру БМ з декількома сотнями вершин, використовуючи навчальну вибірку з мільйонів записів. До таких методів відносяться Tetrad II [3] та SopLeq [4].

3.1 Методи на основі оціночних функцій

Для побудови БМ у вигляді дерева Чу і Ліу (Chow and Liu) в 1968 році запропонували алгоритм, що ґрунтується на використанні значень взаємної інформації між вершинами. Як рішення метод видає структуру із значенням спільного розподілу ймовірностей мережі, яке найбільше відповідає навчальним даним. Побудова структури БМ здійснюється за $O(N^2)$ кроків, де N – кількість вершин мережі. Однак цей алгоритм не працює для багатозв’язаних БМ.

У 1988 році Рібан і Перл (Rebane and Pearl) запропонували удосконалений модифікований алгоритм Чу і Ліу для побудови БМ у вигляді полідерева. Купер і Гершкович (Cooper and Herskovits) в 1990 році розробили алгоритм Кутато (Kutato). На етапі ініціалізації алгоритму вважається, що всі вершини БМ незалежні; після цього обчислюється ентропія цієї мережі. Потім виконується додавання дуг між вершинами у мережі таким чином, щоб мінімізувати ентропію БМ. Для роботи алгоритму потрібна наявність УМВ.

Купер і Гершкович в 1992 році запропонували широко відомий алгоритм К2, який виконує пошук структури з максимальним значенням функції Купера-Гершковича (КГ). Для роботи алгоритму потрібна наявність УМВ. В 1994 році запропоно-

вано алгоритм HGC. Цей алгоритм суттєво відрізняється від інших (що ґрунтуються на оціночних функціях) тим, що вперше саме в ньому були використані два нових поняття: (1) параметричної модульності (parametric modularity) та (2) рівнозначності подій (event equivalence). Інші дослідники досить довго не використовували одночасно цих понять. Але одночасне застосування цих понять дозволяє об'єднувати статистичну інформацію та експертні знання для побудові БМ.

Вонг і Ксіанг (Wong and Xiang) запропонували в 1994 році алгоритм для побудови Марковських мереж з використанням значення ентропії та I-map. Граф G ймовірнісної моделі M називають незалежною картою (independency map, скорочено I-map), якщо з незалежності вершин графа G випливає незалежність моделі M. Цей алгоритм дозволяє представити процес, який моделюється, у вигляді I-map, і у випадку, коли мережа є однозв'язною, гарантовано будується БМ. Разом із Чу (Chu) Ксіанг розробив у 1997 році більш швидкодіючий варіант запропонованого алгоритму.

Алгоритм Лема-Бахуса (Lam-Bacchus), запропонований в 1996 році, виконує евристичну побудову структури мережі, використовуючи значення взаємної інформації між вершинами, а як оціночна функція використовується функція опису мінімальною довжиною (minimum description length).

Алгоритм Бенедикта (Benedict), запропонований в 1996 році, виконує евристичний пошук на основі УМВ, аналізуючи умовні незалежності в структурі мережі на основі *d*-розділення, а як функція оцінки використовується ентропія.

CB алгоритм запропоновано в 1995 році. Він використовує ТУН між вершинами мережі, для побудови ВМВ. Для побудови структури мережі використовується функція КГ.

Алгоритм Фрідмана-Голдшмідта (Friedman-Goldszmidt) запропонований в 1996 році. Для побудови мережі використається аналіз її локальних підструктур, а як оціночна функція використовується функція опису мінімальною довжиною (ОМД) та оцінка Байєса.

В алгоритмі WKD, запропонованому в 1996 році, за оціночну функцію при побудові мережі використано функцію повідомлення мінімальної довжини (minimum message length), яка схожа на ОМД.

Алгоритм Сузукі (Suzuki), запропоновано у 1999 році, заснований на методі гілок та границь для задавання послідовності побудови структури мережі, а як оціночна функція використовується ОМД.

Також існує множина різноманітних поглинаючих алгоритмів (greedy algorithm), в яких для оцінювання можна використовувати різноманітні функції, наприклад максимальної правдоподібності або байєсівський інформаційний критерій.

3.2 Методи на основі використання тестів на умовну незалежність

У 1983 році Вермут і Лоуренс (Wermuth and Lauritzen) запропонували алгоритм для побудови структури БМ, застосовуючи ТУН. Цей алгоритм виконує послідовний перебір УМВ. Для кожної пари вершин X_k та X_t, таких, що $X_t < X_k$ (тобто X_k – це предок для X_t), виконується обчислення значення умовної незалежності. Цей алгоритм гарантує побудову БМ за навчальними даними, але при цьому потрібно обчислити велику кількість ТУН між вершинами, що можливо лише у випадку, коли мережа складається з невеликої кількості вершин.

У 1988 році Перл (Pearl) запропонував алгоритм побудови скінченного спрямованого ациклічного графа (boundary DAG algorithm). Цей алгоритм будує БМ, маючи ВМВ та функцію спільного розподілу (або достатньо велику навчальну вибірку даних). Разом із будь-яким, не досить складним, методом пошуку цей алгоритм позбавлений проблеми, яка полягає у необхідності розрахунку великої кількості тестів на умовну незалежність, застосовуючи алгоритм Вермута і Лоуренса. Однак необхідність обчислення великої кількості ТУН виникає при застосуванні цього алгоритму для побудови марковських мереж, тобто мереж із прихованими вузлами.

У 1990 році запропоновано SRA алгоритм, який є модифікацією алгоритму скінченого спрямованого ациклічного графа. Цей алгоритм висуває менш жорсткі вимоги до упорядкування множини вершин. Для побудови БМ достатньо мати частково упорядковану множину вершин та ще деякі обмеження. Побудова БМ виконується послідовним додаванням дуг між вершинами з використанням евристичного пошуку. Але алгоритм виконує експоненціальну кількість розрахунків тестів на умовну незалежність.

Алгоритм «Конструктор» (constructor algorithm) запропоновано у 1990 році. Він дуже схожий на алгоритм побудови скінченого спрямованого ациклічного графа. Замість БМ виконується спроба побудувати марковську мережу. Відмінність цього методу від інших, які використовують ТУН, полягає у тому, що він не виконує надлишкові тести на умовну незалежність і йому не потрібна упорядкована множина вершин.

Алгоритму SGS, запропонованому у 1990 році, для побудови структури не потрібна наявність УМВ, але замість неї йому доводиться виконувати експоненціальну кількість тестів на умовну незалежність між вершинами.

PC алгоритм, розроблений в 1991 році, представляє собою удосконалений варіант SGS алгоритму. Цей алгоритм розроблено спеціально для побудови розріджених (sparse) БМ, тобто для мереж із невеликою кількістю дуг між вершинами. Алгоритм KDB, запропонований у 1996 році, для визначення напряму побудови мережі використовує значення взаємних ймовірностей. За оціночну функцію використовується функціонал, що мінімізує значення мережі. Алгоритм FBC (full Bayesian network), запропонований в 2006 році, представляє собою удосконалений алгоритм KDB, який як функцію оцінки при побудові мережі використовує функцію сумарних значень ЗВІ вершин.

3.3 Інші методи

Не завжди побудована структура БМ однозначно відповідає процесу, який моделюється. Інколи це пов’язано з неповнотою даних спостережень або недостатньою визначеністю предметної області. Замість побудови однієї найкращої структури БМ деякі алгоритми як результат видають кілька мережних структур.

Іноді дослідник може не мати всієї інформації про процес, який моделюється, тобто деякі змінні, які впливають на процес, відсутні. Їх називають прихованими змінними (hidden variables) або латентними змінними (latent variables). Існують алгоритми *евристичного пошуку* [5], [6], які намагаються враховувати такі приховані змінні при моделюванні.

Для випадку, коли навчальні дані неповні або частина з них невірна (missing data), запропоновано декілька алгоритмів *стиснення границь* (bound and collapse) та група алгоритмів, які використовують значення *максимального математичного очікування* (expectation maximization, або скорочено ЕМ).

Метод стиснення границь [7] моделює відсутність даних, припускаючи, що ймовірність відсутніх даних приймає значення в інтервалі від 0 до 1, тобто виконується аналіз цього інтервалу на відсутність даних за наявною інформацією. Після цього виконується стиснення границь інтервалу в точку шляхом використання опуклої комбінації з точок екстремумів, використовуючи інформацію про неповні дані.

Алгоритм максимізації математичного очікування був запропонований у 1977 році в [8]. Алгоритм намагається знайти локальні оптимальні оцінки максимальної правдоподібності параметрів. Головна ідея алгоритму полягає у тому, що за наявності значень усіх вузлів, навчання (на кроці M) буде простим, оскільки наявна вся необхідна інформація. Тому на кроці E виконується обчислення значення очікуваної правдоподібності (expectation of likelihood), включаючи латентні змінні, так ніби вони спостерігались. На кроці M робиться обчислення значення максимальної правдоподібності параметрів, використовуючи максимізацію значень очікуваної правдоподібності отриманих на кроці E. Далі алгоритм знову виконує крок E з використанням параметрів, отриманих на кроці M , і так далі.

На основі алгоритму максимізації математичного очікування розроблено серію подібних алгоритмів [9], [10]. Так, наприклад, структурний алгоритм максимізації математичного очікування поєднує у собі стандартний алгоритм максимізації математичного очікування, що оптимізує параметри, та алгоритм структурного пошуку моделі відбору. Цей алгоритм будує мережі, ґрунтуючись на штрафних ймовірнісних значеннях, які включають значення, отримані за допомогою байєсівського інформаційного критерію, принципу мінімальної довжини опису, а також значення інших критеріїв.

Висновки

Виконано огляд методів побудови (навчання) структури мереж Байєса. Показано, що на сьогодні існує множина методів структурного навчання МБ та критеріїв оптимізації, які можна використати при їх побудові. Наявність великої кількості методів формування структури МБ свідчить про те, що існують проблеми стосовно розв'язання цієї задачі, які неможливо розв'язати за допомогою одного-двох методів. Це проблеми, пов'язані із високою розмірністю задач, наявністю змінних різних типів та вимогами до якості результату – імовірнісного висновку. Тому вибір методу навчання структури мережі повинен ґрунтуватись на докладному поглибленому аналізі задачі, яка розв'язується за допомогою мережі, та можливості отримання достовірних експертних і статистичних даних. Враховуючи можливу неоднозначність отриманого розв'язку, структуру мережі необхідно будувати за двома-трьома альтернативними методами і вибрати потім кращий розв'язок.

У майбутніх дослідженнях доцільно автоматизувати процес побудови структури мережі за деякою множиною альтернативних методів при розбитті загальної вибірки на навчальну та валідаційні набори даних, включаючи вибір кращої зі структур за критеріями структурної різниці або перехресної ентропії. Це дасть можливість уникнути можливої неоднозначності вибору.

Література

1. Методы и модели анализа данных OLAP и Data Mining / [А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод]. – СПб : БВХ-Петербург, 2004. – 336 с.
2. Чубукова И.А. Data Mining / Чубукова И. А. – М. : Бином ЛБЗ, 2008. – 384 с.

3. Spirtes P. Causation, prediction and search / P. Spirtes, C. Glymour and R. Scheines // Adaptive computation and machine learning, MIT press. – January 2001. – 565 p.
4. Jouffe L. New search strategies for learning Bayesian networks / Jouffe L. and Munteanu P. // Proc. of tenth international symposium on applied stochastic models and data analysis (ASMDA 2001). – Compiegne (France). 12 – 15 June 2001. – Vol. 2. – P. 591-596.
5. Spirtes P. Heuristic greedy search algorithms for latent variable models / P. Spirtes, T. Richardson and C. Meek // Proc. of artificial intelligence and sta структурний алгоритм максимізації математичного очікування tistics (AI & Statistics 1997), Fort Lauderdale (Florida). – 1997. – P. 481-488.
6. Verma T. Equivalence and synthesis of causal models / T. Verma and J. Pearl // Proc. of the sixth international conference on Uncertainty in Artificial Intelligence (UAI'90), Cambridge, Massachusetts, (USA), 27 – 29 July, 1990. – NY. : Elsevier science, 1991. – P. 255-270.
7. Sebastiani P. Bayesian inference with missing data using bound and collapse / P. Sebastiani and M. Ramoni // Journal of Computational and Graphical Statistics. – 2000. – Vol. 9, № 4. – P. 779-800.
8. Dempster A.P. Maximum likelihood from incomplete data via the EM algorithm / A.P. Dempster, N.M. Laird and D.B. Rubin // Journal of the Royal Statistical Society. – 1977. – Vol. 39, № 1. – P. 1-38.
9. Friedman N. The Bayesian structural EM algorithm / Friedman N. // Fourteenth conference on Uncertainty in Artificial Intelligence (UAI'98), Madison, Wisconsin, (USA), 24 – 26 July, 1998. – SF. : Morgan Kaufmann, 1998. – P. 129-138.
10. Zhang Z. Surrogate maximization (minimization) algorithms for AdaBoost and the logistic regression model / Z. Zhang, J. Kwok and D. Yeung // Proc. of the twenty-first international conference on machine learning (ICML 2004). – 2004. – 117 p.

П.И. Бидюк, А.Н. Терентьев, М.М. Коновалюк

Байесовские сети в технологиях интеллектуального анализа данных

Предложен обзор методов построения (обучения) структуры сетей Байеса (СБ). Показано, что на сегодня существует множество методов структурного обучения СБ и критериев оптимизации, которые можно использовать при их построении. Поэтому выбор метода обучения структуры сети должен базироваться на углубленном анализе задачи, которая решается с помощью сети, и возможности получения достоверных экспертных и статистических данных.

P.I. Bidyuk, O.M. Terentyev, M.M. Konovalyuk

Bayesian networks in technologies of intellectual data analysis

A review is proposed of structural learning for Bayesian networks (BN). It is shown that today exists a wide set of structural learning methods for BN as well as optimization criteria that could be used for learning. That is why the selection of a learning method should be based on profound analysis of the problem to be solved by BN and the possibility of obtaining truthful expert and statistical data.

Стаття надійшла до редакції 15.04.2010.